Alternate Path Reasoning in Intelligent Instrument Fault Diagnosis for Gas Chromatography

Kristin L. Adair and Susan I. Hruska
Department of Computer Science
Florida State University
(adair, hruska)@cs.fsu.edu

John W. Elling
Engineering Sciences and Applications - Measurement Technology
Los Alamos National Laboratory
elling@esa.lanl.gov

Abstract

The analysis of soil contaminants contained in hazardous waste sites is a multi-step process involving gathering soil samples, creating gas chromatograms from the samples, and analyzing the chromatograms for evidence of contaminants in the soil sample. Instrument faults occurring during the creation of gas chromatograms may cause the chromatograms to be unanalyzable for contaminants. The automation of instrument fault diagnosis for gas chromatography is addressed using expert network technology.

An expert network development and testing environment has been fielded by a team of researchers at Florida State University. This environment includes tools for building the expert network, aids for automated knowledge acquisition, and a training algorithm. An introduction to the design and implementation of the intelligent system and techniques for ascertaining when multiple reasoning pathways exist for reaching a particular conclusion in the system are presented in this paper.

1 Introduction

In building knowledge-based systems for real problems such as fault analysis in gas chromatography, a variety of methods for reasoning with uncertainty are useful. The reasoning techniques presented combine knowledge gathered through traditional knowledge engineering with experts in the field and data-driven techniques. The underlying form of the system described is a hybrid intelligence system called an *expert network*. This hybrid method incorporates rule-based knowledge, training of weights in the rule-base using artificial neural network type techniques, and adaptive network structure refinement [2] [6] [8]. The system successfully recognizes faulty chromatograms from a variety of features detected in the graph and further diagnoses the most likely cause of instrument failure. This diagnosis is a critical part of the overall task of automating contaminant analysis in soil samples in an automated laboratory setting.

In this paper, we give a brief introduction to the design and implementation of the intelligent system for machine fault diagnosis and present techniques for ascertaining when multiple reasoning pathways exist for reaching a particular conclusion in the system.

2 Gas chromatogram fault diagnosis

The analysis of soil contaminants contained in hazardous waste sites is a multi-step process involving gathering samples, creating gas chromatograms from the samples, and analyzing the chromatograms for signals from contaminants. Faulty chromatograms may be produced due to errors in sample preparation or via a gas chromatography (GC) instrument fault. When performed by humans, the task of detecting faulty chromatograms by eye is costly and prone to misdiagnoses [5]. Automation of this task is the goal of the research reported here.

To detect GC errors, or *faults*, which occur during analysis, experts typically examine the gas chromatogram of a single sample for features, or *symptoms*, indicative of an instrument failure or sample error. If a fault exists, the gas chromatograph data may not be useful. Therefore, it is important that samples produced from a faulty GC system are discovered with a high degree of certainty before the chromatograms are used. Attempts to automate this portion of the analysis task reveal that the process human experts use in determining machine fault is difficult for experts to articulate precisely. Several types of reasoning are used, and a system built to automate this decision-making process must be robust and adaptable.

Formalizing the representation and reasoning process used in automated GC fault diagnoses has progressed through several stages in our fielded system. First, a knowledge table depicting relationships between each symptom and fault was produced. Based on the knowledge table, a hybrid system called an *expert network* was utilized for reasoning using the expert knowledge. The knowledge table was then structurally refined to show how distinct subsets of symptoms may lead to a single fault. Refinement of the knowledge, both in the strength of connections between entities in the table and in the structure of the knowledge, has required the integration of several computational intelligence paradigms with a synthesis of new techniques.

3 Knowledge table

Rule-based systems provide a natural format for representing the high-level reasoning that experts chromatographers use when they analyze a chromatogram for faults. Early work to represent expert knowledge in GC fault diagnosis in a written form is described by Stillman and Lahiri [7]. Their end result was a knowledge table relating symptoms to faults using *True* and *False*. If a symptom is related to a fault, the entry for the symptom-fault pair contains *True*. A *False* entry indicates the opposite relationship. Otherwise, the entry is left blank.

This basic format was also used by an interdisciplinary team comprised of scientists and engineers from Florida State University, Los Alamos National Laboratory, Sandia National Laboratory, Oak Ridge National Laboratory, and Varian Chromatography Systems to create a more extensive knowledge table for GC instruments of interest. The team first identified a set of symptoms and faults related to GC machine fault diagnosis. The table built relates these symptoms and faults to each other using a set of five semantic qualifiers: *Always*, *Usually*, *Sometimes*, *Infrequently*, and *Never*. These qualifiers give the experts more freedom in expressing symptom-fault relationships. Additionally, a blank is left in the knowledge table if no relationship exists between the symptom and fault. A portion of the table for the faults Contaminated Sample, Leaking Syringe, and Column Bleed is presented in Figure 1. Using the knowledge table, one can determine how each individual symptom affects the conclusion of a fault. From Figure 1, it can be seen that for the fault Contaminated Sample Rising Baseline and Peaks Out Of Range appear *Infrequently*, No Peaks *Never* appears, and Sensitivity Change Up, Sensitivity Change Down, Irregular Baseline, High Noise, and Extra Peaks appear *Sometimes*.

The knowledge table created is used to create a system for diagnosing GC faults. The system developed utilizes the *expert network* hybrid technique, and an inference engine described briefly below and in more detail elsewhere [6].

4 Reasoning from the knowledge table

4.1 Expert networks

An expert network is a hybrid of an expert system and an artificial neural network [6]. While seemingly dissimilar, these two techniques complement each other, and the hybrid system is a more powerful tool than either one alone. In expert networks, the nodes and the connections among nodes are not only functional but also have semantic interpretations.

As a basis, the rules of an expert system can be represented as nodes and connections in a digraph structure. Each node represents an assertion from a rule. Rule antecedents are joined to rule consequent nodes by connections with a strength reflecting the certainty associated with the rule. Assertions are represented uniquely as identifiable nodes in the digraph. An example of the translation from a rule in an expert system to an expert network is shown in Figure 2.

The weighted digraph structure forms the architecture of the expert network. The additional features of an expert system are incorporated to create a network useful in reasoning. The combining of evidence in an expert system is represented as the combining function of the expert network. The activation function in the expert network is taken from the firing function of the expert system inference engine.

Extracting explanations of conclusions reached by an artificial neural network can be difficult. However, because of the representation based on expert knowledge, an explanation facility can be built for expert networks. Expert networks can also be shown to be logically equivalent to any expert system. When represented as a network, however, the certainty factors of the expert system can be trained, the architecture can be refined iteratively using data, and in general, the system assumes a robust, adaptive facility that allows it to evolve over time and with changing conditions.

4.2 Expert network for GC fault diagnosis

Using the knowledge represented in the knowledge table created by the team of experts, an expert network was developed [2]. The expert network for GC fault diagnosis consists of four layers of nodes: Symptom, Filter, Combination, and Fault. The translation of the entries for the fault Contaminated Sample in the knowledge table shown in Figure 1 to an expert network is shown in Figure 3.

In the Symptom layer for each fault, there is a single node for each symptom with a non-blank qualifier for the

Symptoms

Fault	Retention Time Shift	Sensitivity Change Up	Sensitivity Change Down	Tailing Peaks	Clipped Peaks	Rounded Peaks	Irregular Baseline	Rising Baseline	High Noise	Negative Dip After Peak	Cannot Zero Baselin	Extra Peaks	No Peaks	Replicate Precision	Peaks Out Of Range
Contaminated Sample		S	S				S	I	S			S	N		I
Leaking Syringe		N	A				N	N			N		I	U	N
Column Bleed		U					S	A			U			U	

Figure 1. Portion of knowledge table

IF

the stain of the organism is gramneg AND
the morphology of the organism is rod AND
the patient is a compromised host

THEN

there is suggestive evidence (0.6) that the
identity of the organism is pseudomonas

TO THE SUBNETWORK

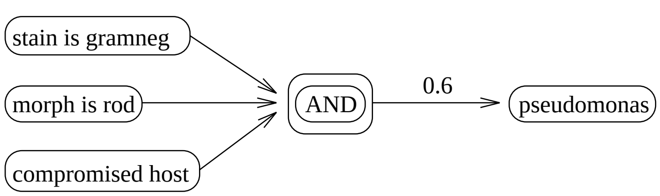

Figure 2. The translation from an expert system rule to an expert network

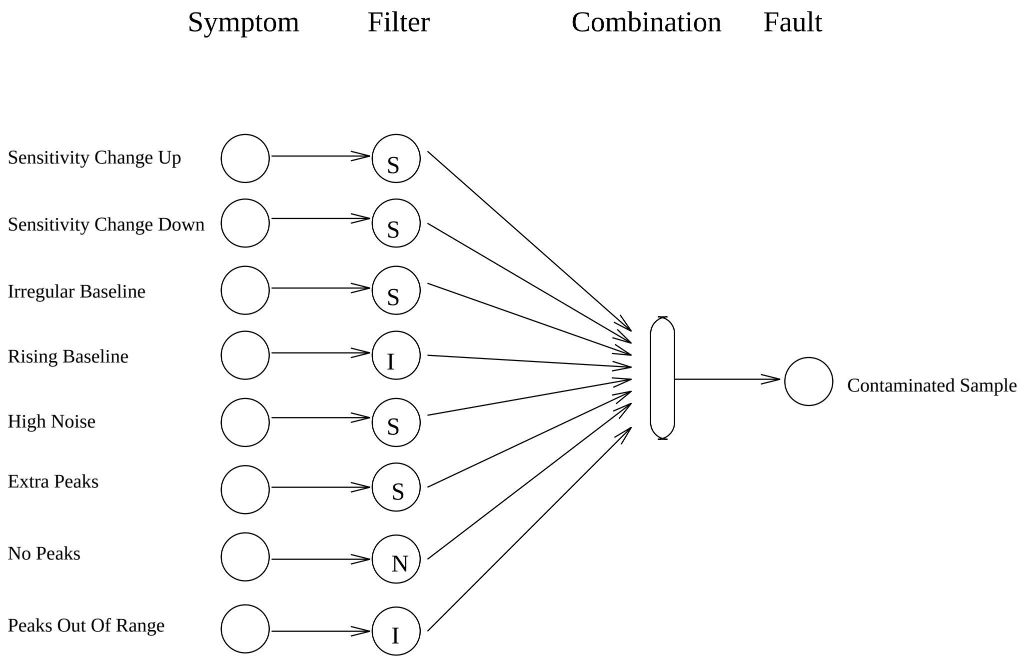

Figure 3. Portion of expert network for Contaminated Sample

fault. Nodes in the Symptom layer are then connected to a node in the Filter layer. There are five types of Filter nodes; one for each semantic qualifier. The Symptom-Filter node pair corresponds to an entry in the knowledge table. Each Filter node is then connected to a Combination node. A Combination node gathers evidence from a set of symptoms for a fault. Each Combination node is then connected to a single Fault node. Each fault in the table is represented by exactly one Fault node.

An additional Fault node is added for diagnosing when no instrument fault has occurred. This “fault,” No Fault, is represented by the symptom set consisting of all symptoms connected to Filter nodes of type Never, corresponding to the belief that a good gas chromatogram will show no symptoms.

A data file for a single sample chromatogram contains the values for the symptom features detected in the gas chromatogram for each symptom in the knowledge table. These incoming values are produced by a signal processing front-end to the expert network [4]. All samples have values for all possible symptoms. Symptom values may range from 0.0 to 1.0. A value of -2.0 represents missing or unavailable data, and its propagation through the network has no effect on the Fault node values.

The symptom values are propagated through their corresponding Symptom nodes to the appropriate Filter node. At the Filter node, the value is multiplied by a positive factor if the symptom should serve as positive evidence for the fault and a negative factor if the symptom serves as negative evidence. The result is then propagated through a weighted connection to a Combination node. At the Combination node, evidence for a symptom set is accumulated [2] and passed to the Fault node. Fault nodes report the value of the highest incoming Combination node. If only one symptom set exists for the fault and therefore only one Combination node, as in Figure 3, the value of the single Combination node is reported. The probable cause of machine failure is the Fault node with the highest value. The resulting network can then be optimized through training procedures. After training the network shows improved diagnostic abilities [2].

5 Multiple paths of reasoning

5.1 Introduction

The knowledge represented in the table formed by the experts relates only how symptoms individually affect a fault. Because no relationships between symptoms were given, each fault was represented by a single symptom set consisting of all symptoms for the fault with a semantic qualifier.

A more realistic view of the knowledge shows that the symptoms connected to a fault in the knowledge table often appear in related groups or subsets, and different subsets of the symptoms appear at different times. For example, a severe case of Contaminated Sample may produce a chromatogram that looks quite different than a milder case. We would like to design the expert network to detect Contaminated Sample at both levels, or anything in between, but this type of knowledge is obscured somewhat in the table by use of semantic qualifiers such as *Sometimes*. While the table could have been further analyzed by human experts, and the training examples further labeled as "Severe Contaminated Sample," "Moderate Contaminated Sample," etc., the manual knowledge acquisition task becomes quickly untenable, with excessive amounts of "intelligent" pre-processing needed before the system would work. Automation of this refinement of the structure of the knowledge led to the following technique for discovering *multiple paths of reasoning* directly from the data [1].

One set of multiple paths discovered from the data is given in Figure 4, and is represented in expert network form in Figure 5. By examining the network, one can see that a path is one way of moving from a set of Symptom nodes to a Fault node. Multiple paths of reasoning mean there are multiple ways of moving from a set of Symptom nodes to a Fault node. Note that the multiple paths are united in a single Fault node for Contaminated Sample; the final evaluation of the likelihood of Contaminated Sample is made by taking the maximum of all values sent by Combination nodes to Contaminated Sample.

Using an ART2 [3] based clustering technique, paths of reasoning are discovered using available symptom data. ART2 is an unsupervised clustering technique that does not require a presumption of the number of categories into which sample data points will fall. To separate a single symptom set into paths of reasoning, an ART2 network is created and trained for each fault. The individual networks cluster all the sample data from a single fault. A probable path of reasoning is then said to exist if two or more clusters of sample data are created for a single fault.

The sample data for a single fault separated into clusters is then used to determine the semantic qualifier entries in the knowledge table for the new alternate paths of reasoning. NetMedic [8], developed by Timpany at Florida State University, analyzes a sample data set to determine the semantic qualifier associated with each symptom, finding the frequency of occurrence for each symptom in a cluster, and then suggesting connections for each symptom in a path. For example, a symptom which has a positive value in all sample data for a cluster will be proposed to have an *Always* connection to that fault. The newly proposed paths are incorporated into the expert network after confirmation from domain experts.

6 Results

Testing of the expert network for machine fault diagnosis illustrates the usefulness of the intelligent system. The network was tested on 667 sample data files taken from chromatograms showing one of twelve possible faults. Before training, the network correctly diagnoses 475 faults (71.2%). Using a training set of 166 randomly chosen samples, the network correctly diagnoses 624 (93.6%) of the complete sample set. The Table 1 shows the results of the expert network with multiple paths of reasoning when trained on all 667 sample data files.

7 Conclusions

Fielding a system to deal with real expert knowledge and real data is a challenge which draws upon a variety of representation and reasoning methods to be effective. In this paper, we consider several methods used in the knowledge engineering phase with experts in gas chromatogram analysis. A knowledge table representing the relationships between symptoms and faults is given and an expert network solution for this table is shown, allowing reasoning based on the knowledge table. The knowledge table representation is expanded to allow for multiple paths of reasoning and the diagnosis of no fault data. Automation of this complex diagnosis task is shown to be approachable using this synthesis of techniques with data-driven discovery complementing the traditional knowledge engineering methods. Extension of this methodology to other diagnosis tasks, especially those with apparently ambiguous knowledge bases and noisy real world data, is evident.

References

[1] K. L. Adair. *Data Driven Optimization of Expert Networks*. PhD thesis, Florida State University, Tallahassee, Florida, Fall 1996.

[2] K. L. Adair, A. P. Levis, and S. I. Hruska. An expert network development environment for automating machine fault diagnosis. In *Applications and Science of Artificial Neural Networks II*, volume 2760 of *SPIE Proceedings*, pages 506–522, 1996.

[3] G. A. Carpenter and S. Grossberg. *Pattern Recognition by Self-Organizing Neural Networks*, chapter ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns, pages 397–423. MIT Press, Cambridge, Massachusetts, 1991.

[4] J. Elling, S. Lahiri, and R. Roberts. Feature extraction algorithms for the qualitative assessment of gas chromatograms. Technical Report LA-UR-95-4157, Los Alamos, October 1995. accpeted for publication in *Journal of Chromatographic Science*.

[5] J. W. Elling, L. K. Klatt, and W. P. Unruh. Automated data interpretation in an automated environmental laboratory. *Laboratory Robotics and Automation*, 6(2):73–78, 1994.

[6] D. C. Kuncicky, S. I. Hruska, and R. Lacher. Hybrid systems: The equivalence of rule-based expert system and artificial neural network inference. *International Journal of Expert Systems*, 4(3):281–297, 1992.

Fault	Retention Time Shift	Sensitivity Change Up	Sensitivity Change Down	Tailing Peaks	Clipped Peaks	Rounded Peaks	Irregular Baseline	Rising Baseline	High Noise	Negative Dip After Peak	Cannot Zero Baselin	Extra Peaks	No Peaks	Replicate Precision	Peaks Out Of Range
Contaminated Sample			S				U	I				I	N		I
Contaminated Sample		I	S				U	U	I			I	N		S

Symptoms

Figure 4. Multiple paths of reasoning in a knowledge table

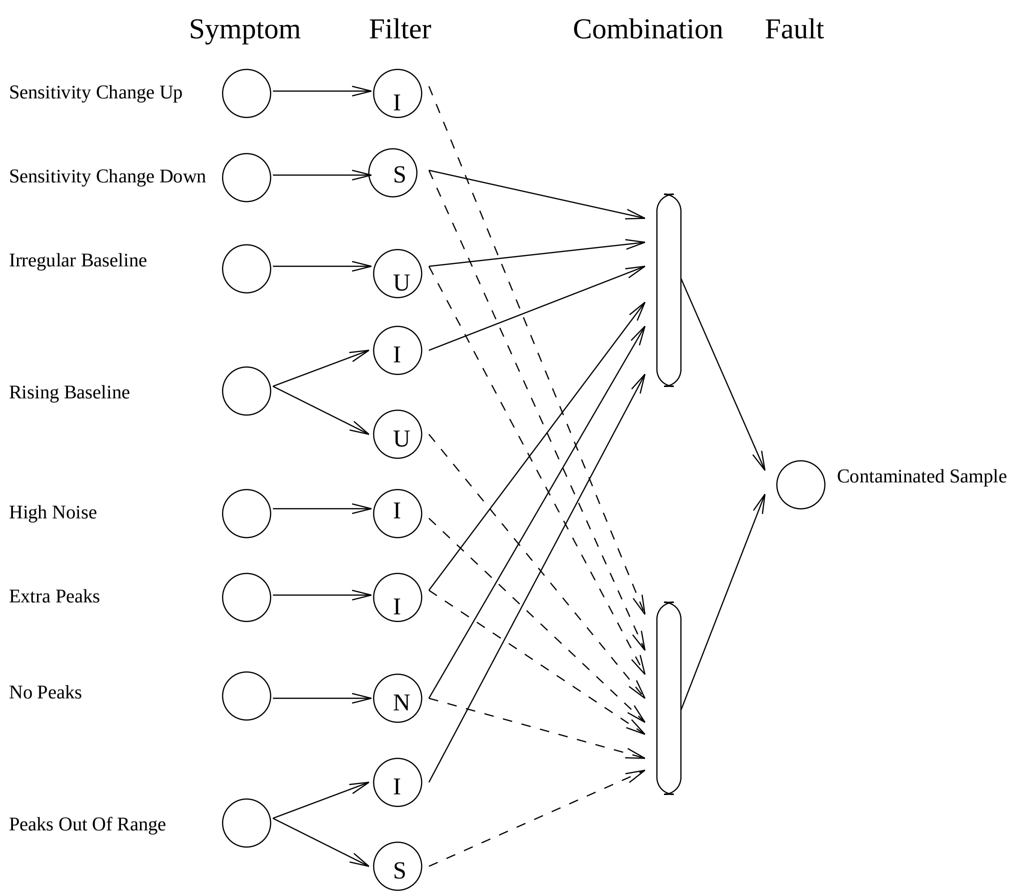

Figure 5. Network with multiple paths of reasoning

Table 1. Performance of expert network with multiple paths of reasoning

Fault	Number of Samples	Number Correct Before Training	Number Correct After Training
Leaking Syringe	66	60	59
Column Bleed	31	23	31
Column Degradation	94	87	93
Makeup Gas Loss	106	62	98
Bad Liner	22	21	18
Leaking Septum	43	23	37
Detector Saturated	6	6	6
Low Carrier Gas	51	9	50
Contaminated Sample	32	23	31
Sample Too Concentrated	43	42	43
No Carrier Gas	47	6	47
No Fault	126	113	125
Total	667	475 71.2%	638 95.7%

[7] S. Lahiri and M. J. Stillman. Expert systems: Diagnosing the cause of problem aas data. *Analytical Chemistry*, 64(4):283–291, February 1992.

[8] R. G. Timpany. Data assisted knowledge acquisition for alteration of table–based expert networks. Master's thesis, Florida State University, Tallahassee, Florida, Summer 1995.